

TITLE
SCORPION TOXINS

5

FIELD OF THE INVENTION

This invention is in the field of molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding scorpion toxins that are sodium channel agonists.

BACKGROUND OF THE INVENTION

10

Alpha neurotoxins are short, single-chain, polypeptides crosslinked by four disulfide bridges, and responsible for insect and mammal poisonings. These neurotoxins show variability in their apparent toxicity, in their primary structures, and in their binding features to neuronal membrane preparations (Dufton and Rochat (1984) *J. Mol. Evol.* 20:120-127). Despite differences in their primary structures and phylogenetic selectivity, scorpion
15 neurotoxins affecting sodium (Na) channels are closely related in their spatial arrangement. And in their compact globular structure kept rigid by the four disulfide bridges (Miranda et al. (1970) *Eur. J. Biochem.* 16:514-523; and Fontecilla-Camps (1989) *J. Mol. Evol.* 29:63-67).

20

Zilbergberg and coworkers determined that single amino acid residues are important for receptor binding and for biological activity of scorpion Na channel toxins (Zilbergberg et al. (1997) *J. Biol. Chem.* 272:14810-14816). As examples, the lysine at position 8 of LqhIT was demonstrated to be necessary for binding activity and toxicity without change in overall structure. A substantial decrease in biological activity without a significant change in structure was found when the aromatic amino acid phenylalanine, at position 17, was
25 substituted for glycine. Conversely, changes in structure are not necessarily associated with differences in toxicity as demonstrated when tyrosine at position 49 was changed to leucine.

25

While potassium (K) channels have been shown to be central to heart function, the role of chlorine- (Cl) and Na-channels in this activity is less clear (Johnson et al.. (1998) *J. Neurogent.* 12:1-24). Sodium entry hyperpolarizes the cell, producing indirect, Na-
30 dependent changes of calcium transport (Friedman (1998) *Annu. Rev. Physiol.* 60:179-197). Abnormal influx of calcium is thought to be very important in the pathogenesis of several central nervous system disorders in vertebrates, including stroke damage, epilepsy, and the neuronal death associated with chronic epilepsy.

30

35

Excitatory amino acids, most notably glutamate and aspartate, are the predominant excitatory neurotransmitter in the vertebrate (including human) central nervous system. These amino acids are released from presynaptic nerve terminals and, after diffusing across the synaptic cleft, contact special receptor molecules in the postsynaptic cell membrane. These receptors indirectly influence the flow of various ions across the cell membrane and

thus contribute to production of an electrical response to the chemical message delivered by neurotransmitter molecules. A number of common and very serious neurological problems involve abnormal function of excitatory amino acid synapses. These include epilepsy, several degenerative disorders such as Huntington's disease, and neuronal death following stroke. Unfortunately, there are very few chemical agents which are potent and selective blockers of excitatory amino acid receptors. Na-channel agonists may be used for these purposes.

A drug with high affinity for the receptor could be expected to produce irreversible blockade of synaptic transmission. When labeled with some tracer molecule, such a drug would provide a reliable way of tagging receptors to permit measurement of their number and distribution within cells and tissues. These features would have very valuable consequences for research on excitatory amino acid neurotransmission and for the development of therapeutic agents to treat central nervous system dysfunction in humans and animals. Methods for treating heart and neurological diseases by applying toxins derived from spiders have been described (U.S. Patent No. 4,925,664).

Arthropod animals, including insects, and certain parasitic worms use excitatory amino acids as a major chemical neurotransmitter at their neuromuscular junction and in their central nervous system. Because of the damage done by insect pests and the prevalence of parasitic worm infections in animals and humans in many countries, there is a constant need for potent and specific new pesticides and anthelmintic drugs that are non-toxic to humans, pets, and farm animals.

Chemical insecticides are an integral component of modern agriculture, and are an effective means for reducing crop damage by controlling insect pests. However, chemical agents are under continuous scrutiny due to the potential for environmental contamination, selection of resistant populations of agronomic pests, and toxicity to non-target organisms such as beneficial insects, aquatic organisms, animals and man. As a result, alternative strategies for insect control are being sought that are effective and yet benign to non-target populations and the environment. One of these strategies is to use microorganisms that are naturally occurring pathogens of target insect populations. The expression of scorpion toxins using baculovirus vectors will be an advantage since these toxins have been previously shown to be highly toxic and very specific (Zlotkin et al. (1995) American Chemical Society, Symposium on Agrochemicals).

Due to a combination of problems associated with some synthetic insecticides, including toxicity, environmental hazards, and loss of efficacy due to resistance, there exists a continuing need for the development of novel means of invertebrate control, including the development of genetically engineered recombinant baculoviruses which express protein toxins capable of incapacitating the host more rapidly than the baculovirus infection per se.

Scorpion venoms have been identified as possible sources of compounds providing insecticidal properties. Two insect-selective toxins isolated from the venom of the scorpion *Leiurus quinquestriatus* and affecting sodium conductance have been reported previously (Zlotkin et al. (1985) *Arch. Biochem. Biophys.* 240:877-87). One toxin, AaIT, induced fast
5 excitatory contractive paralysis of fly larvae and the other, LqhIT2, induced slow depressant flaccid paralysis suggesting that these two toxins have different chemical and pharmacological properties (Zlotkin et al. (1971) *Biochimie (Paris)* 53:1073-1078). Thus, other toxins derived from scorpion venom will also have different chemical and pharmacological properties.

10 SUMMARY OF THE INVENTION

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a first polypeptide of at least 60 amino acids that has at least 95% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a scorpion alpha toxin XIV polypeptide selected from the group
15 consisting of SEQ ID NOs:2, 4, and 6, a scorpion neurotoxin I polypeptide of SEQ ID NO:9, a scorpion depressant toxin LqhIT2 polypeptide selected from the group consisting of SEQ ID NOs:12, 14, and 16. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

It is preferred that the isolated polynucleotides of the claimed invention consist of a
20 nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 8, 11, 13, and 15 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 9, 12, 14, and 16. The present invention also relates to an isolated polynucleotide comprising a nucleotide sequences of at least one of 40 (preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group
25 consisting of SEQ ID NOs:1, 3, 5, 8, 11, 13, 15, and the complement of such nucleotide sequences.

The present invention relates to a chimeric gene comprising an isolated polynucleotide of the present invention operably linked to suitable regulatory sequences.

The present invention relates to an isolated host cell comprising a chimeric gene of the
30 present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as an insect, a yeast or a plant cell, or prokaryotic, such as a bacterial cell or virus. If the host cell is a virus, it is preferably a baculovirus. It is most preferred that the baculovirus comprises an isolated polynucleotide of the present invention or a chimeric gene of the present invention.

35 The present invention relates to a process for producing an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention, the process comprising either transforming or transfecting an isolated compatible host cell with a chimeric gene or isolated polynucleotide of the present invention.

The present invention relates to a scorpion toxin polypeptide selected from the group of alpha toxin XIV, neurotoxin I, and depressant toxin LqhIT2 of at least 60 amino acids comprising at least 95% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 9, 12, 14, and 16.

5 The present invention relates to a method of selecting an isolated polynucleotide that affects the level of expression of a sodium channel agonist polypeptide in a host cell, the method comprising the steps of:

constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention;

10 introducing the isolated polynucleotide or the isolated chimeric gene into a host cell;

measuring the level a alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2 polypeptide in the host cell containing the isolated polynucleotide; and

15 comparing the level of a alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2 polypeptide in the host cell containing the isolated polynucleotide with the level of a alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2 polypeptide in a host cell that does not contain the isolated polynucleotide.

The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2 polypeptide gene, preferably a scorpion alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2 polypeptide gene, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 40 (preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 8, 11, 13, 15, and the complement of such
25 nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of an alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2 amino acid sequence.

30 The present invention also relates to a method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding an alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2 polypeptide comprising the steps of: probing a cDNA or genomic library with an isolated polynucleotide of the present invention; identifying a DNA clone that hybridizes with an isolated polynucleotide of the present invention; isolating the identified DNA clone; and sequencing the cDNA or genomic
35 fragment that comprises the isolated DNA clone.

Another embodiment of the instant invention pertains to a method for expressing a gene encoding a alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2 in the genome of a recombinant baculovirus in insect cell culture or in viable insects wherein said

insect cells or insects have been genetically engineered to express an alpha toxin XIV, a neurotoxin I, or a depressant toxin LqhIT2.

The present invention relates to an expression cassette comprising at least one nucleic acid of Claim 1 operably linked to a promoter.

5 The present invention relates to a method for positive selection of a transformed cell comprising the steps of transforming a plant cell with a chimeric gene of the present invention or an expression cassette of the present invention; and growing the transformed plant cell under conditions allowing expression of the polynucleotide in an amount sufficient to induce insect resistance to provide a positive selection means.

10 BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTING

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

Figure 1 shows a comparison of the amino acid sequences of the alpha toxin XIV of the instant invention (SEQ ID NOs:2, 4 and 6) with the sequence of alpha toxin XIV from 15 *Buthus occitanus* (NCBI General Identifier No. 1041278; SEQ ID NO:7). The conserved cysteine residues probably involved in intrachain disulfide bridges are boxed. The first amino acid of the mature toxin is marked by an arrow above the top row.

Figure 2 shows a comparison of the amino acid sequences of the neurotoxin I of the instant invention (SEQ ID NO:9) with the sequence of neurotoxin I from *Buthus occitanus* 20 *tunetanus* (NCBI General Identifier No. 134335; SEQ ID NO:10). The conserved cysteine residues probably involved in intrachain disulfide bridges are boxed. The first amino acid of the mature toxin is marked by an arrow above the top row.

Figure 3 shows a comparison of the amino acid sequences of the depressant toxin LqhIT2 of the instant invention (SEQ ID NOs:12, 14 and 16) with the sequence of the 25 depressant toxin LqhIT2 from *Leiurus quinquestriatus* (NCBI General Identifier No. 102796; SEQ ID NO:17). The conserved cysteine residues probably involved in intrachain disulfide bridges are boxed. The first amino acid of the mature toxin is marked by an arrow above the top row.

Table 1 lists the polypeptides that are described herein, the designation of the cDNA 30 clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

35

TABLE 1
Scorpion Sodium Channel Agonists

Protein	Clone Designation	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
Scorpion Alpha Toxin XIV	lst.pk0004.e12	1	2
Scorpion Alpha Toxin XIV	lst.pk0016.c5.f	3	4
Scorpion Alpha Toxin XIV	lst.pk0015.h11	5	6
Scorpion Neurotoxin I	lst.pk0013.f1	8	9
Scorpion Depressant Toxin LqhIT2	lst.pk0004.c8	11	12
Scorpion Depressant Toxin LqhIT2	lst.pk0013.c9	13	14
Scorpion Depressant Toxin LqhIT2	lst.pkpk0004.e8	15	16

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, a "polynucleotide" is a nucleotide sequence such as a nucleic acid fragment. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, or synthetic DNA. An isolated polynucleotide of the present invention may include at least one of 40 contiguous nucleotides, preferably at least one of 30 contiguous nucleotides, most preferably one of at least 15 contiguous nucleotides, of the nucleic acid sequence of the SEQ ID NOs:1, 3, 5, 8, 11, 13 and 15.

"NPV" stands for Nuclear Polyhedrosis Virus, a baculovirus. "Polyhedrosis" refers to any of several virus diseases of insect larvae characterized by dissolution of tissues and accumulation of polyhedral granules in the resultant fluid. "PIBs" are polyhedral inclusion bodies. "AcNPV" stands for the wild-type *Autographa californica* Nuclear Polyhedrosis Virus.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the polynucleotide sequence. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not

substantially affect the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or
5 DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms.
10 Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the
15 temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the
20 amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are 95% identical to the amino acid sequences reported herein. Suitable nucleic acid
25 fragments not only have the above homologies but typically encode a polypeptide having at least 50 amino acids, preferably 100 amino acids, more preferably 150 amino acids, still more preferably 200 amino acids, and most preferably 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI).
30 Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

35 A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-

based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without affecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan

appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

“Gene” refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. “Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a nucleotide sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in a variety of cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not

been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol.* 3:225-236).

The "3' non-coding sequences" refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell* 1:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Functional RNA" refers to sense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms.

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

A "signal sequence" is an amino acid sequence that is covalently linked to an amino acid sequence representing a mature protein. The signal sequence directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53).

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides, including signal sequences, present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

It is understood that "an insect cell" refers to one or more insect cells maintained *in vitro* as well as one or more cells found in an intact, living insect.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several scorpion sodium channel agonists have been isolated and identified by comparison of cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other arthropod species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other alpha toxin XIV, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired arthropod employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer

DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding arthropod genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:5673-5677; Loh et al. (1989) *Science* 243:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques* 1:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol.* 36:1-34; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed scorpion sodium channel agonists are expressed. This would be useful as a means for controlling insect pests by producing plants that are more insect-tolerant than the naturally occurring variety.

Expression in plants of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding

sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, LC-MS, or phenotypic analysis.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded scorpion sodium channel agonist. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 8).

Insecticidal baculoviruses have great potential to provide an environmentally benign method for agricultural insect pest control. However, improvements to efficacy are required in order to make these agents competitive with current chemical pest control agents. One approach for making such improvements is through genetic alteration of the virus. For instance, it may be possible to modify the viral genome in order to improve the host range of the virus, to increase the environmental stability and persistence of the virus, or to improve the infectivity and transmission of the virus. In addition, improving the rate at which the virus acts to compromise the infected insect would significantly enhance the attractiveness of insecticidal baculoviruses as adjuncts or replacements for chemical pest control agents. One method for increasing the speed with which the virus affects its insect host is to introduce into the baculovirus foreign genes that encode proteins that are toxic to the insect wherein death or incapacitation of the insect is no longer dependent solely on the course of the viral

infection, but instead is aided by the accumulation of toxic levels of the foreign protein. The results are insecticidal recombinant baculoviruses.

Recombinant baculoviruses expressing the instant scorpion sodium channel agonists (or portions thereof) may be prepared by protocols now known to the art (e.g., Tomalski et al., U.S. Patent No. 5,266,317, exemplifying neurotoxins from the insect-parasitic mites; 5 McCutchen et al. (1991) *Bio/Technology* 9:848-852; Maeda et al. (1991) *Virology* 184:777-780, illustrating construction of a recombinant baculovirus expressing AaIT; also see O'Reilly et al. (1992) *Baculovirus Expression Vectors: A Laboratory Manual*, W. H. Freeman and Company, New York; King and Possee (1992) *The Baculovirus Expression* 10 *System*, Chapman and Hall, London; U.S. Patent No. 4,745,051). These methods of gene expression provide economical preparation of foreign proteins in a eukaryotic expression vector system, in many instances yielding proteins that have achieved their proper tertiary conformation and formed the proper disulfide bridges necessary for activity.

Commonly, the introduction of heterologous genes into the baculovirus genome occurs 15 by homologous recombination between viral genomic DNA and a suitable "transfer vector" containing the heterologous gene of interest. These transfer vectors are generally plasmid DNAs that are capable of autonomous replication in bacterial hosts, affording facile genetic manipulation. Baculovirus transfer vectors also contain a genetic cassette comprising a region of the viral genome that has been modified to include the following features (listed in 20 the 5' to 3' direction): 1) viral DNA comprising the 5' region of a non-essential genomic region; 2) a viral promoter; 3) one or more DNA sequences encoding restriction enzyme sites facilitating insertion of heterologous DNA sequences; 4) a transcriptional termination sequence; and 5) viral DNA comprising the 3' region of a non-essential genomic region. A heterologous gene of interest is inserted into the transfer vector at the restriction site 25 downstream of the viral promoter. The resulting cassette comprises a chimeric gene wherein the heterologous gene is under the transcriptional control of the viral promoter and transcription termination sequences present on the transfer vector. Moreover, this chimeric gene is flanked by viral DNA sequences that facilitate homologous recombination at a non-essential region of the viral genome. Recombinant viruses are created by co-transfecting 30 insect cells that are capable of supporting viral replication with viral genomic DNA and the recombinant transfer vector. Homologous recombination between the flanking viral DNA sequences present on the transfer vector and the homologous sequences on the viral genomic DNA takes place and results in insertion of the chimeric gene into a region of the viral genome that does not disrupt an essential viral function. The infectious recombinant virion 35 consists of the recombined genomic DNA, referred to as the baculovirus expression vector, surrounded by a protein coat.

In a preferred embodiment, the non-essential region of the viral genome that is present on the transfer vector comprises the region of the viral DNA responsible for polyhedrin

production. Most preferred is a transfer vector that contains the entire polyhedrin gene between the flanking sequences that are involved in homologous recombination. Recombination with genomic DNA from viruses that are defective in polyhedrin production (due to a defect in the genomic copy of the polyhedrin gene) will result in restoration of the polyhedrin-positive phenotype. This strategy facilitates identification and selection of recombinant viruses.

In another embodiment, baculoviral genomic DNA can be directly modified by introduction of a unique restriction enzyme recognition sequence into a non-essential region of the viral genome. A chimeric gene comprising the heterologous gene to be expressed by the recombinant virus and operably linked to regulatory sequences capable of directing gene expression in baculovirus-infected insect cells, can be constructed and inserted directly into the viral genome at the unique restriction site. This strategy eliminates both the need for construction of transfer vectors and reliance on homologous recombination for generation of recombinant viruses. This technology is described by Ernst et al. (Ernst et al. (1994) *Nuc. Acid Res.* 22: 2855-2856), and in WO94/28114.

Recombinant baculovirus expression vectors suitable for delivering genetically encoded insect-specific neurotoxins require optimal toxin gene expression for maximum efficacy. A number of strategies can be used by the skilled artisan to design and prepare recombinant baculoviruses wherein toxin gene expression results in sufficient quantities of toxin produced at appropriate times during infection in a functional form and available for binding to target cells within the insect host.

The isolated toxin gene fragment may be digested with appropriate enzymes and may be inserted into the pTZ-18R plasmid (Pharmacia, Piscataway, NJ) at the multiple cloning site using standard molecular cloning techniques. Following transformation of *E. coli* DH5 α MCR, isolated colonies may be chosen and plasmid DNA prepared. Positive clones will be identified and sequenced with the commercially available forward and reverse primers.

Spodoptera frugiperda cells (Sf-9) may be propagated in ExCell[®] 401 media (JRH Biosciences, Lenexa, KS) supplemented with 3.0% fetal bovine serum. Lipofectin[®] (50 μ L at 0.1 mg/mL, Gibco/BRL) may be added to a 50 μ L aliquot of the transfer vector containing the toxin gene of interest (500 ng) and linearized polyhedrin-negative AcNPV (2.5 μ g, Baculogold[®] viral DNA, Pharmigen, San Diego, CA). Sf-9 cells (approximate 50% monolayer) may be co-transfected with the viral DNA/transfer vector solution. The supernatant fluid from the co-transfection experiment may be collected at 5 days post-transfection and recombinant viruses may be isolated employing standard plaque purification protocols, wherein only polyhedrin-positive plaques will be selected (Granados, R. R., Lawler, K. A., *Virology* (1981), 108, 297-308).

To propagate the recombinant virus of interest, isolated plaques may be picked and suspended in 500 μ L of ExCell[®] media supplemented with 2.5% fetal bovine serum. Sf-9 cells in 35 mM petri dishes (50% monolayer) may be inoculated with 100 μ L of the viral suspension, and supernatant fluids collected at 5 days post infection. These supernatant fluids will be used to inoculate cultures for large scale propagation of recombinant viruses.

Expression of the encoded toxin gene by the recombinant baculovirus will be confirmed using a bioassay, LCMS, or antibodies. The presence of toxin activity in the recombinant viruses will be monitored *in vivo*. These assays involve comparison of biological activity of recombinant viruses to wild-type. Third instar larvae of *H. virescens* are infected orally by consumption of diet that contains test and control viruses and the larvae monitored for behavioral changes and mortality.

Isolated plugs of a standard insect diet are inoculated with approximately 5000 PIBs of each virus. Individual larvae that have not fed for 12 h prior to beginning of the bioassay are allowed to consume the diet for 24 h. The larvae are transferred to individual wells in a diet tray and monitored for symptoms and mortality on a daily basis (Zlotkin et al. (1991) *Biochimie (Paris)* 53:1073-1078).

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries: Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various *Leiurus* scorpion telson tissues were prepared. cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP[™] XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP[™] XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via

polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science* 252:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

ESTs encoding scorpion sodium channel agonists were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet.* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Alpha Toxin XIV

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to alpha toxin XIV from *Buthus occitanus* (NCBI General Identifier No. 1041278). Shown in Table 3 are the BLASTP results for individual ESTs:

TABLE 3

BLAST Results for Sequences Encoding Polypeptides Homologous to Alpha Toxin XIV

Clone	BLAST pLog Score 1041278
lst.pk0004.e12	24.15
lst.pk0016.c5.f	28.00
lst.pk0015.h11.f	29.70

The nucleotide sequences from the clones presented above encode entire toxins and all or part of the corresponding signal sequence. The amino acid sequence set forth in SEQ ID NO:2 contains a signal sequence (amino acids 1-11) and a mature toxin (amino acids 12-75). The amino acid sequence set forth in SEQ ID NO:4 contains a signal sequence (amino acids 1-12) and a mature toxin (amino acids 13-79). The amino acid sequence set forth in SEQ ID NO:6 contains a signal sequence (amino 1-19) and a mature toxin (amino acids 20-87).

Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:2, 4 and 6 and the *Buthus occitanus* sequence (NCBI General Identifier No. 1041278). The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4 and 6 and the *Buthus occitanus* sequence (SEQ ID NO:7).

TABLE 4

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Alpha Toxin XIV

SEQ ID NO.	Percent Identity to 1041278
2	60.0
4	68.4
6	64.7

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*, 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode three distinct, entire, scorpion alpha toxin XIV, two of which have partial signal sequences and one has the entire signal sequence.

EXAMPLE 4

Characterization of cDNA Clones Encoding Neurotoxin I

The BLASTX search using the EST sequence from clone 1st.pk0013.f1 revealed similarity of the protein encoded by the cDNAs to neurotoxin I from *Buthus occitanus tunetanus* (NCBI General Identifier No. 134335), with a pLog value of 36.70. The amino acid sequence set forth in SEQ ID NO:9 contains a signal sequence (amino acids 1-19) and a mature protein (amino acids 20-84). Figure 2 presents an alignment of the amino acid sequences set forth in SEQ ID NO:9 and the *Buthus occitanus* sequence (SEQ ID NO:10).

The amino acid sequence presented in SEQ ID NO:9 is 80.0% identical to the *Buthus occitanus* sequence.

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5). Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragment encodes an entire scorpion neurotoxin I with its signal sequence.

EXAMPLE 5

Characterization of cDNA Clones Encoding Depressant Toxin LqhIT2

The BLASTX search using the EST sequences from clones listed in Table 5 revealed similarity of the polypeptides encoded by the cDNAs to depressant toxin LqhIT2 from *Leiurus quinquestriatus* (NCBI General Identifier No. 102796). Shown in Table 5 are the BLAST results for individual ESTs:

TABLE 5

BLAST Results for Sequences Encoding Polypeptides Homologous to Depressant Toxin LQHIT2

Clone	BLAST pLog Score 102796
lst.pk0004.c8	39.75
lst.pk0013.c9	32.40
lst.pkpk0004.e8	18.85

The nucleotide sequences from the clones presented above encode entire toxins and all or part of the corresponding signal sequence. The amino acid sequence set forth in SEQ ID NO:12 contains a signal sequence (amino acids 1-21) and a mature protein (amino acids 22-85). The amino acid sequence set forth in SEQ ID NO:14 contains a signal sequence (amino acids 1-21) and a mature protein (amino acids 22-85). The amino acid sequence set forth in SEQ ID NO:6 contains a signal sequence (amino acids 1-19) and a mature protein (amino acids 20-85).

Figure 3 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:12, 14 and 16 and the *Leiurus quinquestriatus* sequence (NCBI General Identifier No. 102796). The data in Table 6 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:12, 14 and 16 and the *Leiurus quinquestriatus* sequence (SEQ ID NO:17).

TABLE 6

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Depressant Toxin LqhIT2

SEQ ID NO.	Percent Identity to 102796
12	86.9
14	72.1
16	45.9

- 5 Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for
10 pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode three distinct entire scorpion depressant toxin LqhIT2 proteins with their signal sequences.

EXAMPLE 615 Expression of Chimeric Genes in Monocot Cells

- A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain
20 reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (Nco I or Sma I) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes Nco I and Sma I and fractionated on an agarose gel. The
25 appropriate band can be isolated from the gel and combined with a 4.9 kb Nco I-Sma I fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sal I-Nco I promoter fragment of the maize 27 kD zein
30 gene and a 0.96 kb Sma I-Sal I fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence

analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

5 The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu
10 et al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

15 The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat*
20 gene in p35S/Ac is under the control of the 35S promoter from cauliflower mosaic virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

 The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm
25 in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant
30 removed. The particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad
35 Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

 For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of

about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology* 8:833-839).

EXAMPLE 7

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and

placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

5 Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70, U.S. Patent
10 No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from cauliflower mosaic virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225
15 (from *E. coli*; Gritz et al.(1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

20 To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL CaCl₂ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can
25 be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally
30 bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh
35 media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into

individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

5

EXAMPLE 8

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by

centrifugation and re-suspended in 50 μ L of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant
5 determined. One μ g of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 9

Expression of Chimeric Genes in Insect Cells

10 The cDNAs encoding the instant polypeptides may be introduced into the baculovirus genome itself. For this purpose the cDNAs may be placed under the control of the polyhedron promoter, the IE1 promoter, or any other one of the baculovirus promoters. The cDNA, together with appropriate leader sequences is then inserted into a baculovirus transfer vector using standard molecular cloning techniques. Following transformation of *E. coli*
15 DH5 α , isolated colonies are chosen and plasmid DNA is prepared and is analyzed by restriction enzyme analysis. Colonies containing the appropriate fragment are isolated, propagated, and plasmid DNA is prepared for cotransfection.

Spodoptera frugiperda cells (Sf-9) are propagated in ExCell[®] 401 media (JRH Biosciences, Lenexa, KS) supplemented with 3.0% fetal bovine serum. Lipofectin[®]
20 (50 μ L at 0.1 mg/mL, Gibco/BRL) is added to a 50 μ L aliquot of the transfer vector containing the toxin gene (500 ng) and linearized polyhedrin-negative AcNPV (2.5 μ g, Baculogold[®] viral DNA, Pharmingen, San Diego, CA). Sf-9 cells (approximate 50% monolayer) are co-transfected with the viral DNA/transfer vector solution. The supernatant fluid from the co-transfection experiment is collected at 5 days post-transfection and
25 recombinant viruses are isolated employing standard plaque purification protocols, wherein only polyhedrin-positive plaques are selected (O'Reilly et al. (1992), *Baculovirus Expression Vectors: A Laboratory Manual*, W. H. Freeman and Company, New York.). Sf-9 cells in 35 mM petri dishes (50% monolayer) are inoculated with 100 μ L of a serial dilution of the viral suspension, and supernatant fluids are collected at 5 days post infection. In order to
30 prepare larger quantities of virus for characterization, these supernatant fluids are used to inoculate larger tissue cultures for large scale propagation of recombinant viruses. Expression of the instant polypeptides encoded by the recombinant baculovirus is confirmed by bioassay.